Current Biology, Vol. 15, 347–352, February 22, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.cub.2005.02.023

Identification of Transcriptional Regulatory Elements in Chemosensory Receptor Genes by Probabilistic Segmentation

Steven A. McCarroll,[1] Hao Li,[2] and Cornelia I. Bargmann[1,2,3,*]
[1]Department of Anatomy
[2]Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143
[3]Howard Hughes Medical Institute
The Rockefeller University
1230 York Avenue
New York, New York 10021

Summary

Genome sequencing has allowed many gene regulatory elements to be identified through cross-species comparisons [1–5]. However, the expression of genes in multigene families can diverge rapidly between related species [6–9]. An alternative approach to characterizing multigene families utilizes the fact that genes within the group are likely to share aspects of their regulation. Here, we use a statistical approach, probabilistic segmentation [10], to identify sequences that are overrepresented in the regions upstream of *C. elegans* chemosensory receptor genes. Although each of these elements is present in only a subset of the genes, their distribution across and within the promoters of chemosensory receptor genes makes it possible to detect them. Many of the motifs show positional preference with respect to the translational start site and correspond to the binding sites of known families of transcription factors. We verified one motif, the E-box sequence WWYCACSTGYY, by showing that it directs expression of reporter genes to the ADL chemosensory neurons. Thus, probabilistic segmentation can be used to identify functional regulatory elements with no previous knowledge of gene expression or regulation. This approach may be of particular value for rapidly evolving genes in the immune system and the nervous system.

Results and Discussion

Segmentation of *C. elegans* Chemoreceptor Promoter Sequences into 404 Candidate Motifs

We identified potential regulatory motifs for the large chemosensory receptor gene families in *C. elegans* by probabilistic segmentation, a method based on the identification of short DNA sequences, or "words," that are statistically overrepresented in a set of sequences [10]. Probabilistic segmentation makes efficient use of information that is dispersed across a large number of genes while making minimal assumptions about how regulatory elements are distributed across those genes. A sequence data set was generated with 1 kb upstream of the predicted start sites of 921 likely chemosensory receptor genes of the *sra, srb, src, srd, sre, srh, sri, srj, srm, srn, sro, srp, srr, srs, sru, srv, srw, srx*, and *str* families, as predicted by Hugh Robertson (http://www.wormbase.org/) [11, 12]. This 920 kb of sequence was optimally segmented into short sequences or words by the MobyDick implementation of probabilistic segmentation [10] (see Supplemental Data available online with this article). In the optimal segmentation, almost 60% of the promoter sequence was segmented into one-letter words and more than 90% was segmented into words of length five or less (Table 1). About 8% of the sequence was segmented into 404 words of six or more nucleotides, which collectively appeared 7,345 times, or about 20 times each (Table 1 and Table S1).

These 404 long words had several features suggesting that they represent nonrandom regulatory elements. Most known transcriptional control elements can appear on either the coding or the noncoding DNA strand. Among the 404 motifs identified by probabilistic segmentation, there were 35 pairs of inverse complements (versus fewer than two pairs expected by chance, $p < 10^{-20}$). In addition, 71 of these 404 long words fell into families of related sequences that differed at only one nucleotide or that shared a common six-nucleotide core.

Positional and Functional Specificity of Candidate Motifs

Transcriptional control elements are statistically enriched within 200 nt of transcriptional start sites [13, 14], and some regulatory elements have even stronger positional preferences [15]. Twelve of the 404 candidate motifs showed positional preference with respect to the nearest translational start site at $p < 10^{-4}$ by chi-square test ($p < 0.1$ after Bonferroni correction) (Table 2); all twelve showed strong preference for the proximal 200 nt of the promoter region. A more sensitive statistical test based on the binomial distribution identified nine additional motifs that were overrepresented in the proximal 200 nt of sequence (Supplemental Data and Table 2). Most of these motifs corresponded to known binding sites for families of transcription factors.

Motifs with an E-Box Core

Twelve of the 21 motifs shared the E-box core sequence CASCTG on either the coding or noncoding strand. E-boxes (CANNTG) are bound by transcription factors of the basic helix-loop-helix family; specificity for particular family members is determined by the two interior nucleotides (NN) and by nucleotides flanking the E-box core. The frequencies of the core E-box sequences CACCTG, CAGGTG, and CAGCTG in *C. elegans* chemoreceptor promoters all peaked between −40 and −120 (Figure 1A). By contrast, the similar E-box sequence CACGTG (which did not appear in the probabilistic segmentation results) did not show any positional preference within the chemoreceptor gene family (Figure 1A).

SMAD Binding Motifs

Two motifs, GTCTAG and CTAGAC, are complementary sequences with a common positional preference, sug-

*Correspondence: cori@rockefeller.edu

Table 1. Distribution of Word Lengths and Their Occurrences in a Probabilistic Segmentation of the Upstream Sequences of 921 *C. elegans* Chemoreceptor Genes

Length	Words in Dictionary	Appearances in Segmentation	Sequence Explained
1	4	356053	57.54%
2	3	14674	4.74%
3	17	20438	9.91%
4	75	15694	10.14%
5	90	12019	9.71%
6	86	4347	4.21%
7	74	1425	1.61%
8	128	1145	1.48%
9	77	304	0.44%
10	31	92	0.15%
11	8	32	0.06%

gesting that the motif they identify can appear on either the coding or the noncoding strand. The frequency of these motifs was greatest at positions between −40 and −180 (Figure 1B). In mice, this sequence is bound by transcription factors of the SMAD family [16].

CdxA Binding Sequence

The CTATAATT motif showed a positional preference that peaked between −60 and −120; the motif also showed a strand preference, with more appearances at almost every position than its reverse complement sequence AATTATAG, which was not identified in the probabilistic segmentation (Figure 1C). CTATAATT has been identified experimentally as a binding site for the mammalian caudal-type homeobox domain transcription factor CdxA [17], whose orthologs in *C. elegans* include *pal-1, lin-39*, and *ceh-13*. E-box, SMAD, and CdxA motifs typically appeared only once per chemoreceptor gene promoter.

If these sequence motifs represent elements dedicated to the chemosensory system, then we would expect them to be overrepresented among chemosensory

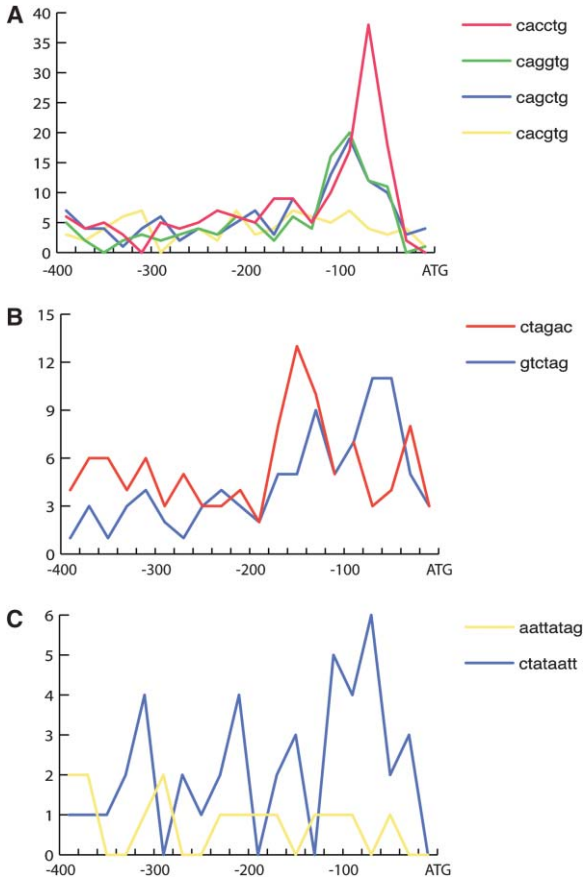

Figure 1. Positional Preference of Candidate Regulatory Motifs and Related Sequences

(A) Positional preference in *C. elegans* chemosensory receptor gene promoters of the four core E-box motifs. CACGTG was not identified by probabilistic segmentation.

(B) Positional preference in *C. elegans* chemosensory receptor promoters of SMAD binding motifs.

(C) Positional preference and strand bias of the CdxA binding motif and its inverse complement sequence.

Table 2. Sequence Motifs with Statistically Significant Positional Preference Upstream of *C. elegans* Chemoreceptor Genes

Word	−1000 to −900	−900 to −800	−800 to −700	−700 to −600	−600 to −500	−500 to −400	−400 to −300	−300 to −200	−200 to −100	−100 to ATG	P_{binom}	P_{chisq}	Comment
TCACCTG	9	4	7	5	6	4	6	13	*20	*43	10^{-15}	10^{-19}	E-box
CAGGTGAAA	1	2	2	1	4	1	1	2	7	*24	10^{-13}	10^{-17}	E-box
CACCTGC	6	3	4	4	3	2	5	7	*16	*33	10^{-15}	10^{-17}	E-box
ACAGGTGAA	1	1	1	0	0	1	0	3	5	*13	10^{-9}	10^{-9}	E-box
AGCAGCTGAAA	0	0	1	0	0	1	0	0	0	*8	10^{-6}	10^{-8}	E-box
GCAGGTGAA	0	4	0	1	2	0	1	0	5	*12	10^{-8}	10^{-7}	E-box
GTCTAG	12	10	20	13	13	10	11	15	*26	*37	10^{-8}	10^{-5}	SMAD
GCAGGTG	6	7	3	2	9	9	4	4	7	*22	10^{-8}	10^{-5}	E-box
TCCACCTGTT	0	0	0	1	0	1	0	0	0	*6	10^{-5}	10^{-5}	E-box
CACCTGTC	2	1	5	0	1	1	1	4	*7	*11	10^{-6}	10^{-4}	E-box
CTAGAC	12	12	21	16	14	13	22	15	*37	*27	10^{-5}	10^{-4}	SMAD
CTATAATT	1	6	4	2	5	12	8	8	10	*17	10^{-4}	10^{-3}	CdxA
CATTTTTC	15	17	19	20	26	18	29	25	27	*45	10^{-5}	10^{-3}	
GTAGACA	7	7	3	8	6	10	2	5	*19	8	10^{-4}	10^{-3}	
CAGCTG	22	24	26	26	25	26	23	18	*39	*48	10^{-6}	10^{-3}	E-box
TGAGCTTT	3	1	4	4	6	8	3	3	*14	*10	10^{-5}	10^{-2}	
CAATAACA	3	4	5	0	9	6	3	4	10	*12	10^{-4}	10^{-2}	
GACAGGTG	1	2	1	3	0	2	1	1	3	*8	10^{-4}	10^{-2}	E-box
GAGCAGCTGA	0	0	0	0	1	0	0	1	1	*4	10^{-4}	NS	E-box
CTTCCTTGT	0	0	0	1	0	0	0	0	*3	2	10^{-3}	NS	
TTACTCAT	6	6	3	9	3	6	5	6	*14	*11	10^{-4}	NS	

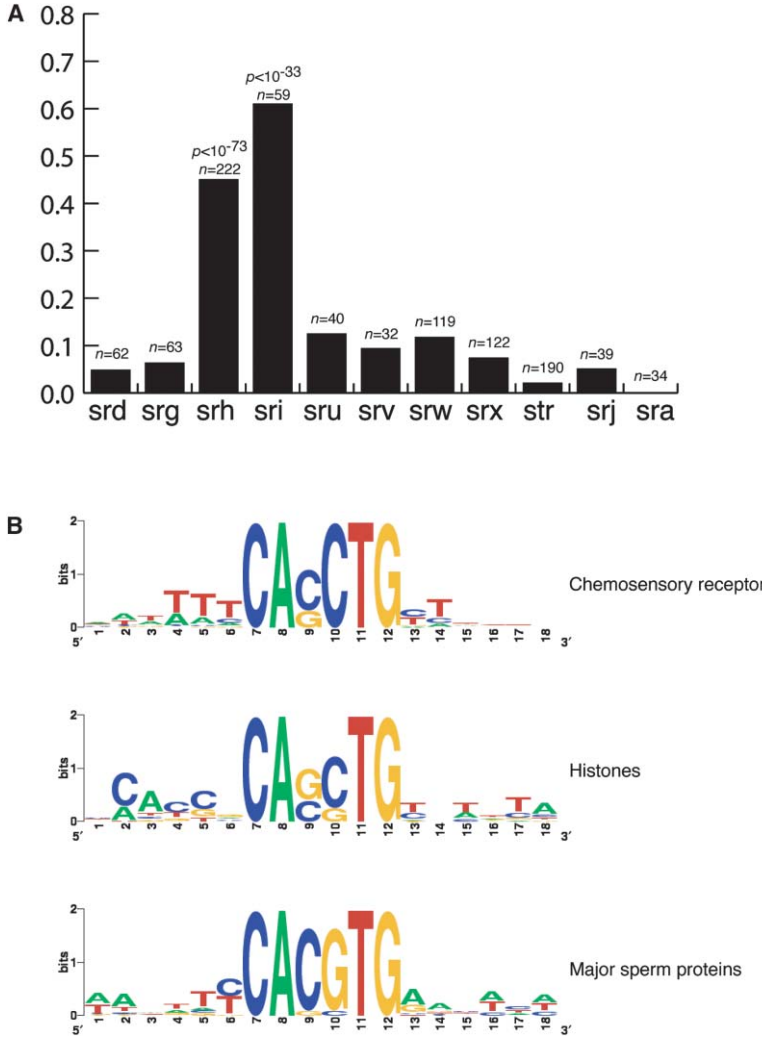

Figure 2. Distribution of CASCTG Motif across Chemosensory Receptor Subfamilies

(A) Fraction of genes in a chemoreceptor subfamily with CASCTG E-box motifs in the proximal promoter. n, total number of genes in the family. The motif appeared in the proximal promoter of 36 of 59 *sri* genes (61%) and 100 of 222 *srh* genes (45%) versus a genome-wide frequency of 4.4%.

(B) Extended E-box motifs associated with chemoreceptor genes, histone genes, and major sperm protein genes. Sequence logos were generated with the WebLogo program [28].

genes relative to their frequency in all genes. To investigate this hypothesis, we identified all occurrences of the E-box motif, the SMAD motif, and the CdxA binding motif in the proximal promoters of all predicted *C. elegans* genes and asked if each motif was statistically overrepresented in any Gene Ontology (GO) categories—about 600 categories of genes defined by common molecular functions, subcellular localization, or biological roles [18]. The E-box core sequences CACCTG (on either strand) and CAGCTG were overrepresented in the proximal promoter regions of two functional sets of genes: G protein coupled receptors and histones (Table S2). By contrast, the E-box motif CACGTG, which was not identified by probabilistic segmentation analysis, was overrepresented in cell motility genes, appearing in the proximal promoters of most genes with major sperm protein (MSP) domains; we suggest that the CACGTG motif may represent a promoter element used to drive gene expression in sperm. The candidate SMAD binding motif and the candidate CdxA motif were both overrepresented specifically in the proximal promoters of G protein coupled receptors. These three motifs thus appear to show high functional specificity.

E-Box and SMAD Motifs Are Differentially Distributed among Chemoreceptor Subfamilies

The *C. elegans* chemoreceptor genes have been classified into multiple families, some of which appear to have independent evolutionary origins [11, 12, 19]. E-box sequences were strongly overrepresented in the *srh* and *sri* families relative to other chemoreceptor subfamilies (Figure 2A). The *srh* gene family is massively expanded in *C. elegans* relative to the related nematode *C. briggsae* [11, 20], but the E-box enrichment was not simply caused by recent gene duplication in the *srh* and *sri* gene families: E-box sequences were more broadly shared than any other 6-mer in the *srh* and *sri* promoters, and within these families, there was no significant relationship between the homology of coding sequences and the likelihood that promoter E-box sequences were shared.

The SMAD motif was overrepresented in genes of the *str* family, appearing in 26 of 190 genes (14% versus a background frequency in the genome of 3.2%; $p < 10^{-9}$). *C. elegans* SMADs are targets of a TGFß pathway that regulates entry into and exit from the alternative dauer larva stage [21, 22]. Dauer formation regulates the ex-

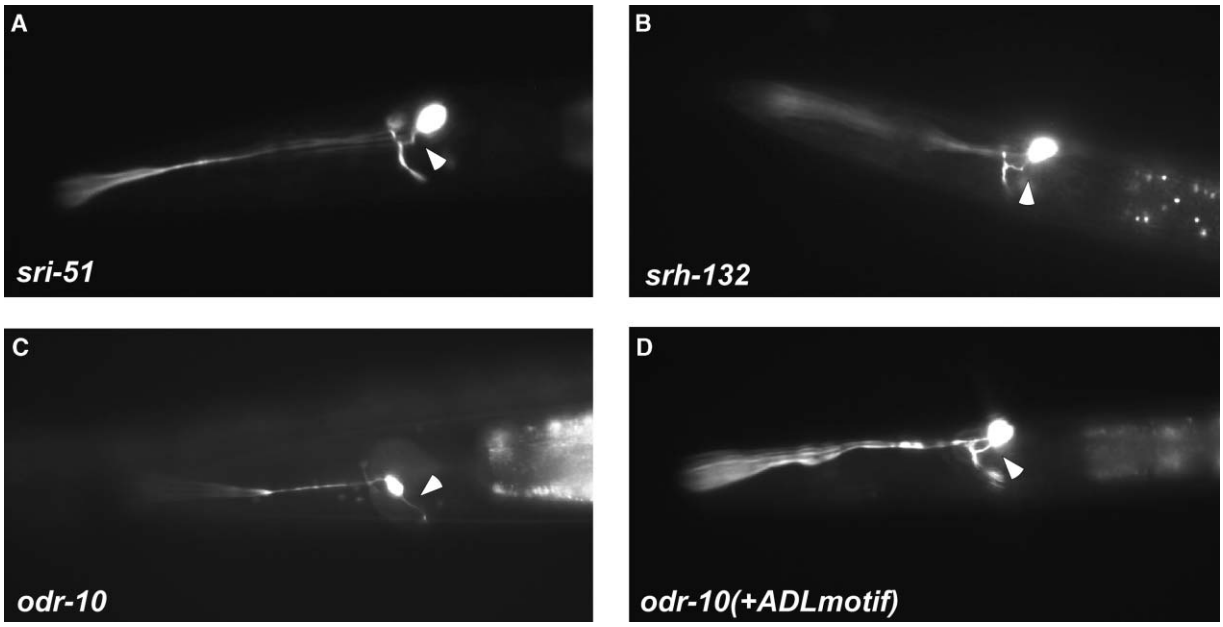

Figure 3. Expression of Transgenes with the Extended E-Box ADL Motif

(A) *sri-51::GFP* transgene expressed in ADL neuron.
(B) *srh-132::GFP* transgene expressed in ADL neuron. This transgene is also expressed in the head mesodermal cell (not shown).
(C) *odr-10::GFP* transgene expressed in AWA neuron.
(D) *odr-10(+ADL motif)::GFP* transgene expressed in ADL neuron. Anterior is at left, and ventral is down. Arrowheads denote characteristic anterior extension of the ADL axon (A, B, and D) and ventral extension of the AWA axon (C) from the cell body. The sensory axons extend a short distance to the nerve ring. The dendrites of all sensory neurons extend anteriorly. Autofluorescence of the gut is visible in (B), (C), and (D).

pression of the *str-2* and *str-3* chemosensory receptor genes [23, 24], both of which have the SMAD motif in their proximal promoters. The CdxA motif was randomly distributed among chemoreceptor subfamilies.

Three Extended E-Box Motifs Can Be Defined by Sequence Context

To examine the immediate sequence context in which candidate transcriptional motifs occur, we analyzed the flanking sequence around each E-box, SMAD, or CdxA motif in chemosensory receptor genes (see Supplemental Data). Appearance of the asymmetric E-box core sequence CACCTG on either DNA strand in chemoreceptor promoter sequences strongly biased the flanking nucleotide composition, suggesting the larger motif WWYCACCTGYY (Table S3 and Figure 2B). This WWY CASCTGYY motif describes 15 of the 404 words identified in the original probabilistic segmentation. Unlike the E-box core, the CdxA motif and the SMAD motif did not appear to be part of larger consensus sequences.

Table 3. Genes of Known or Predicted Function with Strong Matches to the ADL Motif Sequence WWYCASCTGYY in Their Proximal Promoters with Position of Motif and Match to Consensus Sequence

Gene	Motif Position	Motif Score	Gene Description
srh-132	−58	8.36	Chemoreceptor (this work)
srh-186	−70	8.69	Chemoreceptor (this work)
sri-51	−104	8.34	Chemoreceptor (this work)
srh-220	−96	8.43	Chemoreceptor
sro-1	−66	7.91	Chemoreceptor
hlh-2	−127	8.28	bHLH transcription factor (da ortholog)
osm-9	−249	8.10	TRPV cation channel
gpa-1	−341	8.31	G protein
nlp-7	−199	8.05	Neuropeptide
nlp-10	−145	8.23	Neuropeptide
cam-1	−228	8.06	Receptor tyrosine kinase
tax-6	−69	8.03	Calcineurin A
acy-2	−93	8.12	Adenylyl cyclase
C18E3.6	−109	8.32	Adenylyl cyclase-associated protein
C50B6.11	−149	7.97	Neurotransmitter-gated ion channel
ina-1	−60	8.26	Integrin alpha chain
ZK721.4	−147	8.08	Adenosine A3 receptor

Genes known to be expressed in ADL are shown in bold (expression patterns from http://www.wormbase.org/).

A similar analysis of CASCTG E-boxes in the proximal promoters of 29 histone genes yielded the distinct consensus sequence CAYSRCASSTG, whereas CACGTG E-boxes in the proximal promoters of 17 major sperm protein yielded a third consensus sequence, TYCACGT GRA (Figure 2B). The different consensus sequences for the E-boxes in these three gene families suggest that they are bound by distinct transcription factors.

The Extended E-Box Motif WWYCASCTGYY Appears in ADL-Expressed Genes and Acts as an ADL Enhancer Element

To better understand the biological relevance of these computational results, we characterized the E-box motif WWYCASCTGYY. Two chemoreceptor genes with strong matches to proximal WWYCASCTGYY consensus, *srh-220* and *sro-1*, have previously been characterized [19, 24] (Y. Zhang and C.I.B., unpublished data); both were expressed in the ADL chemosensory neuron pair. We tested three additional genes with a strong proximal match to the consensus. *srh-132::GFP, srh-186::GFP*, and *sri-51::GFP* transgenes were all expressed robustly in ADL (three transgenic lines each; Figure 3, Table S4, and data not shown). These results suggest that the extended E-box sequence is associated with expression in ADL neurons.

To assess whether the consensus WWYCACSTGYY sequence is sufficient for ADL expression, we inserted the sequence TTTCACCTGTT into the proximal promoter of the *odr-10* chemosensory receptor gene, which is expressed in the AWA neurons [25]. Animals carrying a *odr-10(+ADL motif)::GFP* transgene showed bright GFP expression in ADL in addition to expression in AWA (Figure 3). The TTTCACCTGTT sequence also enhanced ADL expression when introduced into a *ttx-3* reporter gene fusion but did not induce efficient ADL expression when inserted into the gut-specific *elt-2* promoter (data not shown). Similarly, a 16-nucleotide AIY neuron motif functions in neuron-specific promoters, but not in muscle- or gut-specific promoters [4].

Several other genes in the *C. elegans* genome have the strong matches to the ADL motif in their proximal promoter sequences (Table 3). These known and candidate ADL-expressed genes encode many proteins with neuronal functions, including ion channels, neuropeptides, adenylyl cyclase signaling, and G protein-signaling pathways. The E-box motif is probably not the only route to ADL expression, however; some known ADL-expressed genes lack the motif, and deletion of the motif in the *srh-220* promoter reduced but did not abolish expression in ADL (data not shown).

Conclusions

Using a statistical algorithm, we identified an 11-nucleotide E-box motif associated with expression in a single chemosensory cell type, the ADL neuron. Insertion of the 11-nucleotide ADL motif into the promoter of a *C. elegans* chemosensory receptor gene normally expressed in AWA neurons was sufficient for expression in ADL. Like a 16-nucleotide motif that specifies expression in AIY interneurons [4], this ADL motif appears to be associated with a particular neuronal identity. Other candidate motifs in Table 2 and Table S1 seem likely to be transcriptional regulatory elements as well.

The simplicity of the ADL motif may contribute to evolvability of *Caenorhabditis* chemosensory behaviors: the appearance or disappearance of this sequence in chemosensory receptor promoters could easily alter receptor expression and thereby the behavioral responses to particular odors.

The presence of an ADL motif in approximately half of the promoters in the *srh* and *sri* chemoreceptor gene subfamilies might reflect the use of ADL to sense a particular class of ligands. ADL senses some volatile repellents [19], but it mediates repulsion only when animals have been starved, not when they are well fed [26]. On food, ADL stimulates aggregation of *C. elegans* into social groups [27]. The *srh* gene family is massively expanded in *C. elegans* compared to the closely related nematode *C. briggsae* [20]. A further characterization of ADL function may reveal features of the specific ecological niche of *C. elegans*.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures and four tables and can be found with this article online at http://www.current-biology.com/cgi/content/full/15/4/347/DC1/.

Acknowledgments

We thank Eric Siggia, Hugh Robertson, Jim Thomas, and Oliver Hobert for discussions and comments on this manuscript and Yongmei Zhang for characterizing the *srh-220* promoter. This work was supported by a Howard Hughes Predoctoral Fellowship to S.A.M., by National Institutes of Health DC04089 (C.I.B.), by National Institutes of Health GM70808 (H.L.), and by a Packard Fellowship (H.L.). C.I.B. is an Investigator of the Howard Hughes Medical Institute.

Received: November 23, 2004
Revised: December 19, 2004
Accepted: December 21, 2004
Published: February 22, 2005

References

1. Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science *288*, 136–140.
2. Zheng, P., Pennacchio, L.A., Le Goff, W., Rubin, E.M., and Smith, J.D. (2004). Identification of a novel enhancer of brain expression near the apoE gene cluster by comparative genomics. Biochim. Biophys. Acta *1676*, 41–50.
3. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.
4. Wenick, A.S., and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. Dev. Cell *6*, 757–770.
5. Bigelow, H.R., Wenick, A.S., Wong, A., and Hobert, O. (2004). CisOrtho: a program pipeline for genome-wide identification of transcription factor target genes using phylogenetic footprinting. BMC Bioinformatics *5*, 27.
6. Zylka, M.J., Dong, X., Southwell, A.L., and Anderson, D.J. (2003). Atypical expansion in mice of the sensory neuron-specific Mrg G protein-coupled receptor family. Proc. Natl. Acad. Sci. USA *100*, 10043–10048.
7. Herrada, G., and Dulac, C. (1997). A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. Cell *90*, 763–773.
8. Insel, T.R., and Shapiro, L.E. (1992). Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. Proc. Natl. Acad. Sci. USA *89*, 5981–5985.

9. Young, L.J., Huot, B., Nilsen, R., Wang, Z., and Insel, T.R. (1996). Species differences in central oxytocin receptor gene expression: comparative analysis of promoter sequences. J. Neuroendocrinol. *8*, 777–783.
10. Bussemaker, H.J., Li, H., and Siggia, E.D. (2000). Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. Proc. Natl. Acad. Sci. USA *97*, 10096–10100.
11. Robertson, H.M. (2000). The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. Genome Res. *10*, 192–203.
12. Robertson, H.M. (1998). Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. *8*, 449–463.
13. Haggerty, T.J., Zeller, K.I., Osthus, R.C., Wonsey, D.R., and Dang, C.V. (2003). A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. Proc. Natl. Acad. Sci. USA *100*, 5313–5318.
14. Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat. Genet. *28*, 327–334.
15. Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103–113.
16. Norwitz, E.R., Xu, S., Jeong, K.H., Bedecarrats, G.Y., Winebrenner, L.D., Chin, W.W., and Kaiser, U.B. (2002). Activin A augments GnRH-mediated transcriptional activation of the mouse GnRH receptor gene. Endocrinology *143*, 985–997.
17. Margalit, Y., Yarus, S., Shapira, E., Gruenbaum, Y., and Fainsod, A. (1993). Isolation and characterization of target sequences of the chicken CdxA homeobox gene. Nucleic Acids Res. *21*, 4915–4922.
18. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.
19. Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., and Bargmann, C.I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell *83*, 207–218.
20. Stewart, M.K., Clark, N.L., Merrihew, G., Galloway, E.M., and Thomas, J.H. (2004). High genetic diversity in the chemoreceptor superfamily of *C. elegans*. Genetics, in press. Published online November 1, 2004. 10.1534/genetics.104.035329
21. Daniels, S.A., Ailion, M., Thomas, J.H., and Sengupta, P. (2000). egl-4 acts through a transforming growth factor-beta/SMAD pathway in *Caenorhabditis elegans* to regulate multiple neuronal circuits in response to sensory cues. Genetics *156*, 123–141.
22. Inoue, T., and Thomas, J.H. (2000). Targets of TGF-beta signaling in *Caenorhabditis elegans* dauer formation. Dev. Biol. *217*, 192–204.
23. Peckol, E.L., Troemel, E.R., and Bargmann, C.I. (2001). Sensory experience and sensory activity regulate chemosensory receptor gene expression in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA *98*, 11032–11038.
24. Nolan, K.M., Sarafi-Reinach, T.R., Horne, J.G., Saffer, A.M., and Sengupta, P. (2002). The DAF-7 TGF-beta signaling pathway regulates chemosensory receptor gene expression in *C. elegans*. Genes Dev. *16*, 3061–3073.
25. Sengupta, P., Chou, J.C., and Bargmann, C.I. (1996). *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. Cell *84*, 899–909.
26. Chao, M.Y., Komatsu, H., Fukuto, H.S., Dionne, H.M., and Hart, A.C. (2004). Feeding status and serotonin rapidly and reversibly modulate a *Caenorhabditis elegans* chemosensory circuit. Proc. Natl. Acad. Sci. USA *101*, 15512–15517.
27. de Bono, M., Tobin, D.M., Davis, M.W., Avery, L., and Bargmann, C.I. (2002). Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli. Nature *419*, 899–903.
28. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.